Computational Methods for Supply Chain Disruption Forecasting: A Comparative Analysis

Jennifer Y. Wu

PLIANT Institute, Palo Alto, CA

Journal of Computational Social Science, Vol. 7, No. 4 (December 2024), pp. 891–934

Abstract

Supply chain disruption forecasting has attracted substantial methodological attention over the past decade, producing a fragmented literature in which competing approaches are rarely evaluated against common benchmarks or real-world operational requirements. This paper provides a systematic comparative analysis of five computational approaches to disruption forecasting — statistical time-series models, gradient-boosted ensembles, deep learning architectures, agent-based simulation, and hybrid methods combining statistical and machine learning components — evaluated against a common dataset of corridor-level disruption events across 340 international logistics corridors over six years. Performance is assessed across four dimensions: predictive accuracy at 30, 60, and 90-day horizons; computational efficiency; interpretability for operational deployment; and robustness to data quality degradation. Results indicate that gradient-boosted ensemble methods achieve the strongest balance across all four dimensions for the operational requirements typical of logistics continuity planning, while deep learning architectures offer superior accuracy at short horizons when data quality is high. Implications for model selection in operational deployment contexts are discussed.

Keywords: disruption forecasting, machine learning, gradient boosting, deep learning, supply chain, computational methods, logistics corridors

1. Introduction

The problem of supply chain disruption forecasting — predicting when and where disruptions will occur across complex logistics networks, at time horizons relevant to operational decision-making — has attracted methodological attention from operations research, machine learning, and computational social science communities over the past decade. The resulting literature is rich but fragmented: competing approaches are evaluated against different datasets, at different time horizons, using different performance metrics, and with different assumptions about the operational deployment context in which forecasts will be used.

This fragmentation makes it difficult for practitioners — logistics operators, government continuity planners, and research institutions — to make informed model selection decisions for operational deployment. A practitioner seeking to implement a disruption forecasting capability faces a literature that offers strong claims for multiple competing approaches without the comparative evidence needed to evaluate those claims against a consistent set of operational requirements.

This paper addresses the comparative gap by evaluating five computational approaches against a common dataset and a common set of operational requirements. The evaluation draws on the PLIANT Atlas Intelligence Database disruption event records — the most comprehensive longitudinal dataset of corridor-level disruption events currently available to the research community — and on the operational requirements of logistics continuity planning as articulated through PLIANT's government and industry partner engagements.

The paper makes three contributions. First, it provides the most comprehensive head-to-head comparison of disruption forecasting methods published to date, covering five methodological families across a common six-year dataset. Second, it introduces an operational requirements framework for model evaluation that goes beyond predictive accuracy to incorporate computational efficiency, interpretability, and robustness — dimensions that are critical for real-world deployment but rarely assessed in the academic literature. Third, it provides practical guidance for model selection across different deployment contexts, which has been absent from the existing literature.

2. Methodological Families

2.1 Statistical Time-Series Models

Statistical time-series approaches to disruption forecasting include ARIMA variants, exponential smoothing methods, and structural time-series models that decompose observed disruption patterns into trend, seasonal, and irregular components. These approaches are well-established in the operations management literature and offer strong interpretability and computational efficiency. Their primary limitation for disruption forecasting is the assumption of stationarity — that the statistical properties of the series are constant over time — which is violated when geopolitical dynamics or structural changes in logistics networks alter the underlying disruption generating process.

2.2 Gradient-Boosted Ensemble Methods

Gradient-boosted ensemble methods — most prominently XGBoost and LightGBM — have emerged as a leading approach in applied machine learning for tabular data, offering strong predictive performance across a wide range of feature types with relatively modest data requirements. For disruption forecasting, these methods can incorporate the heterogeneous feature set — geopolitical event indicators, port throughput anomalies, weather data, trade flow statistics — that characterizes logistics corridor data without the preprocessing requirements of deep learning architectures. SHAP values and other model-agnostic explanation methods provide interpretability tools that support operational deployment.

2.3 Deep Learning Architectures

Deep learning approaches to disruption forecasting include recurrent neural networks, transformer architectures, and temporal convolutional networks designed to capture long-range temporal dependencies in sequential data. These approaches have demonstrated strong performance on short-horizon forecasting tasks where data quality is high and the feature-outcome relationship is complex. Their primary limitations for operational deployment are computational cost, data requirements, and interpretability — characteristics that are challenging in the logistics forecasting context where data quality is variable and operational transparency is valued.

2.4 Agent-Based Simulation

Agent-based simulation models disruption dynamics by representing the behavior of individual actors — shipping lines, terminal operators, freight forwarders, customs authorities — and simulating how their interactions produce corridor-level disruption patterns under different scenario conditions. This approach is distinctive in its capacity to model counterfactual scenarios and to represent the adaptive behavior of actors in response to disruption events. Its primary limitations are computational cost and the challenge of calibrating agent behavioral rules to real-world data.

2.5 Hybrid Methods

Hybrid methods combine statistical and machine learning components to leverage the strengths of each — using statistical models to decompose and preprocess time-series data before applying machine learning methods to the residuals, or using machine learning methods to generate features that are then incorporated into statistical forecasting frameworks. These approaches have shown promise in adjacent forecasting domains but have received limited systematic evaluation in the logistics disruption context.

3. Data and Evaluation Framework

The common dataset used for comparative evaluation covers 340 international logistics corridors across all primary trade routes over the period January 2018 through December 2023. Disruption events are drawn from the PLIANT Atlas Intelligence Database disruption incident record, supplemented by AIS vessel tracking data, GDELT geopolitical event data, and NOAA weather records. The dataset was developed in close collaboration with Dr. S.L. Chen, whose corridor-level vulnerability assessment methodology provided the structural foundation for the disruption event classification scheme used across all five model evaluations.

Model performance is evaluated across four dimensions. Predictive accuracy is assessed using AUC at 30, 60, and 90-day forecast horizons on a held-out 2023 test set. Computational efficiency is assessed as mean inference time per corridor-forecast on standardized hardware. Interpretability is assessed using a structured rubric developed through consultation with PLIANT's government partner operational staff. Robustness is assessed by evaluating performance degradation under three data quality degradation scenarios: missing geopolitical event data, missing proprietary throughput data, and increased label noise in the disruption event classification.

4. Results

Gradient-boosted ensemble methods achieve the strongest overall performance across the four evaluation dimensions, with 30-day AUC of 0.83, mean inference time of 0.3 seconds per corridor-forecast, high interpretability scores, and moderate robustness to data quality degradation. Deep learning architectures achieve superior 30-day AUC (0.87) under high data quality conditions but degrade substantially under the missing data scenarios (30-day AUC 0.71 under missing geopolitical event data) and score poorly on interpretability.

Statistical time-series models perform well on computational efficiency and interpretability but show substantially lower predictive accuracy (30-day AUC 0.64) and high sensitivity to the non-stationarity induced by geopolitical regime changes. Agent-based simulation produces the most scenario-flexible outputs but requires 340 times the inference time of gradient-boosted methods per corridor-forecast, limiting operational deployment feasibility. Hybrid methods achieve moderate performance across all four dimensions without excelling on any single dimension.

4.1 Regional Performance Variation

Performance varies meaningfully across regional corridor groups. Asia-Pacific corridors achieve the strongest accuracy across all methods, reflecting both data quality — the Atlas APAC coverage built through Dr. Chen's field survey program is the deepest in the dataset — and the relative stability of structural disruption drivers in the region. Middle East and North Africa corridors present the most challenging forecasting environment, with all methods showing substantially degraded performance relative to their global averages.

5. Model Selection Guidance

Based on the comparative evaluation, three deployment context profiles warrant distinct model selection recommendations. For operational deployment contexts requiring real-time forecasting across large corridor portfolios with variable data quality — the typical requirement for logistics continuity planning —

gradient-boosted ensemble methods offer the strongest overall performance profile. For research contexts where data quality is high, forecast horizons are short, and interpretability requirements are limited, deep learning architectures offer superior accuracy. For scenario planning contexts requiring counterfactual analysis and actor-behavior modeling, agent-based simulation offers capabilities that the other approaches cannot replicate despite its computational cost.

6. Conclusion

The comparative analysis presented here provides the most systematic evaluation of computational disruption forecasting methods published to date. The results support clear model selection guidance for operational deployment contexts while identifying the conditions under which alternative approaches offer competitive advantages. The Atlas dataset underlying this evaluation will be made available to qualified research partners through PLIANT's credentialed data access program, supporting replication and extension of the comparative analysis as new methods are developed.

References

Baryannis, G., Validi, S., Dani, S., & Antoniou, G. (2019). Supply Chain Risk Management and Artificial Intelligence. *International Journal of Production Research*, 57(7), 2179–2202.

Chen, S. L., Patel, R., & Ferreira, I. (2025). Predictive Disruption Modeling in Cross-Border Logistics: A Machine Learning Approach. *International Journal of Logistics Research and Applications*, 28(1), 112–139.

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*, 29(5), 1189–1232.

Ivanov, D. (2020). Predicting the Impacts of Epidemic Outbreaks on Global Supply Chains. *Transportation Research Part E*, 136, 101922.

Lundberg, S. M., & Lee, S. I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, 30.

PLIANT Institute. (2025). *Predictive Models for Supply Chain Disruption Forecasting*. PLIANT Journal, May 2026.

Sodhi, M. S., Son, B. G., & Tang, C. S. (2012). Researchers' Perspectives on Supply Chain Risk Management. *Production and Operations Management*, 21(1), 1–13.

Vaswani, A., et al. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems*, 30.

Wu, J. Y., & Patterson, J. (2024). Supply Chain Risk Management in Non-Linear Environments. *PLIANT Research Series*, Vol. 8.

Zhao, K., Zuo, Z., & Blackhurst, J. V. (2019). Modelling Supply Chain Adaptation for Disruptions. *International Journal of Production Economics*, 218, 128–141.

Acknowledgments: The author thanks Dr. S.L. Chen, whose Atlas APAC corridor assessment work is the empirical foundation of the regional performance analysis in Section 4.1, and whose collaboration on the dataset development made this comparative analysis possible. Dr. J. Patterson and Dr. M. Torres provided feedback on earlier drafts. Research was conducted under PLIANT internal funding.

Correspondence: j.wu@projectpliant.com